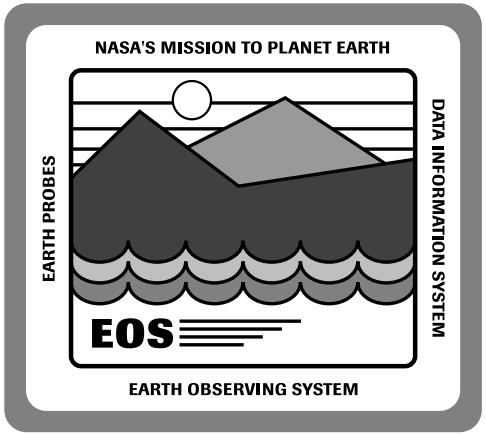
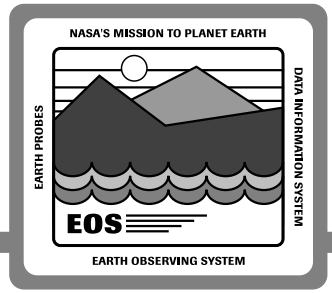


Data Management

Ron Williamson

13 - 14 December 1993

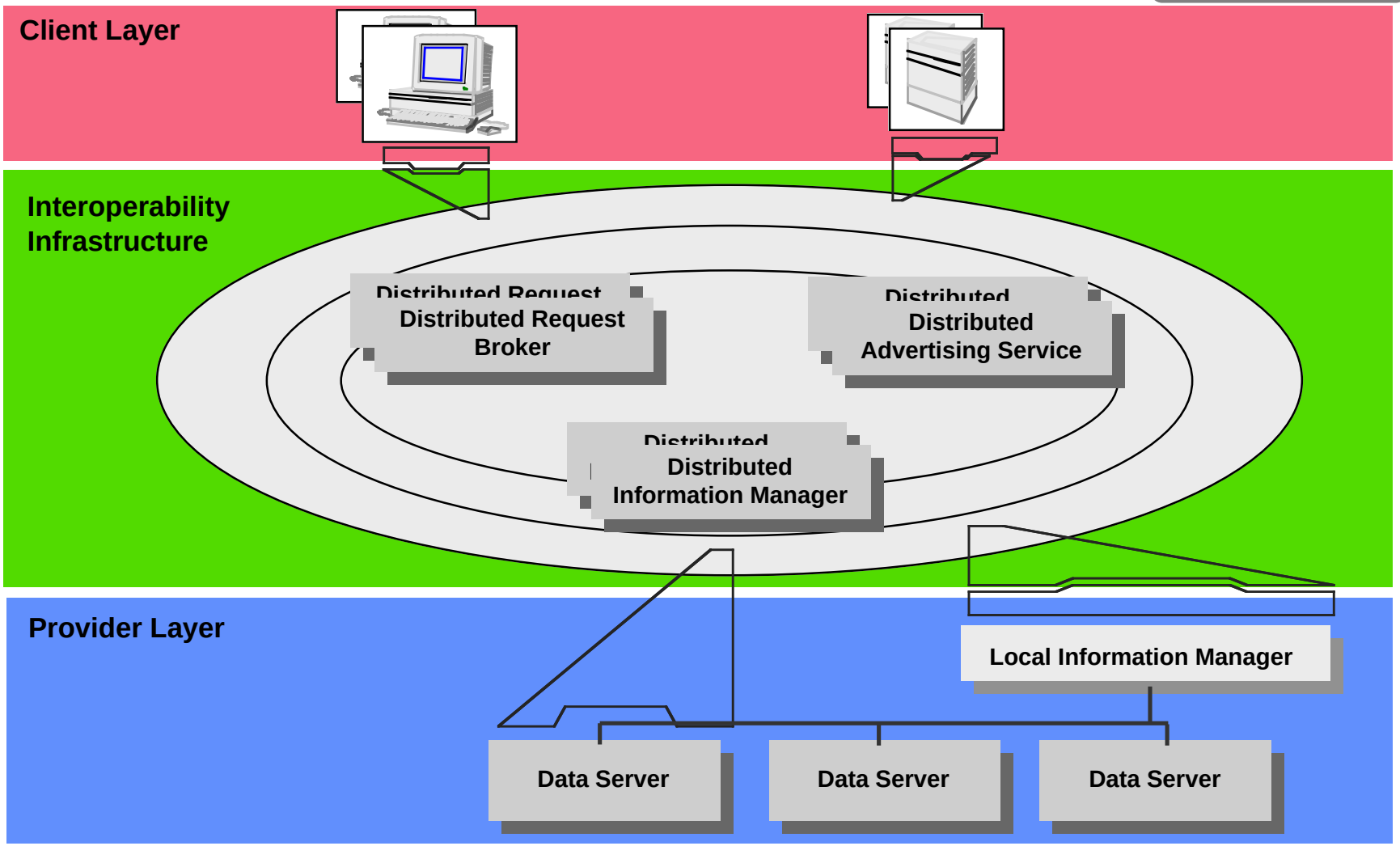
User Needs & Issues



Present New Architecture Through Considering Key User Needs

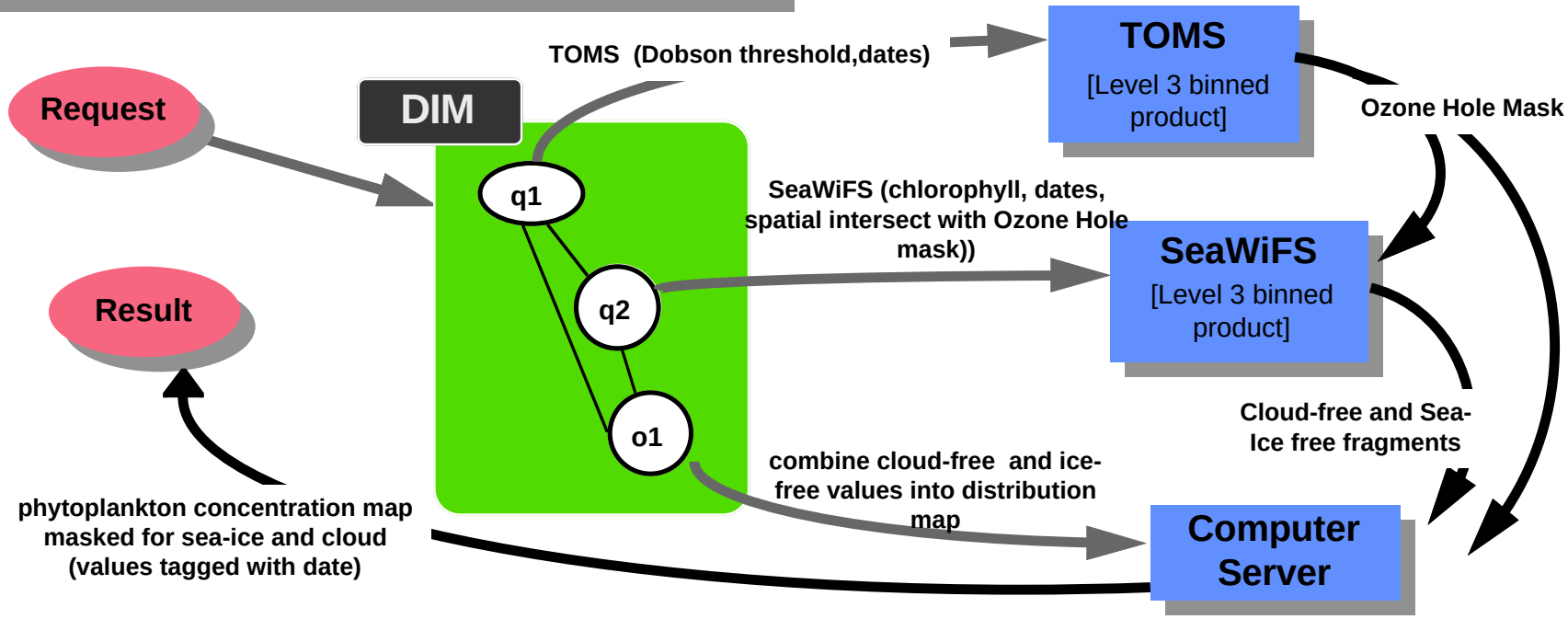
- **Inter-site Searching, including coincidence searching**
- **Content-based Searching**
- **User Access to Data (Request Processing Services)**

Overall Data Management Architecture



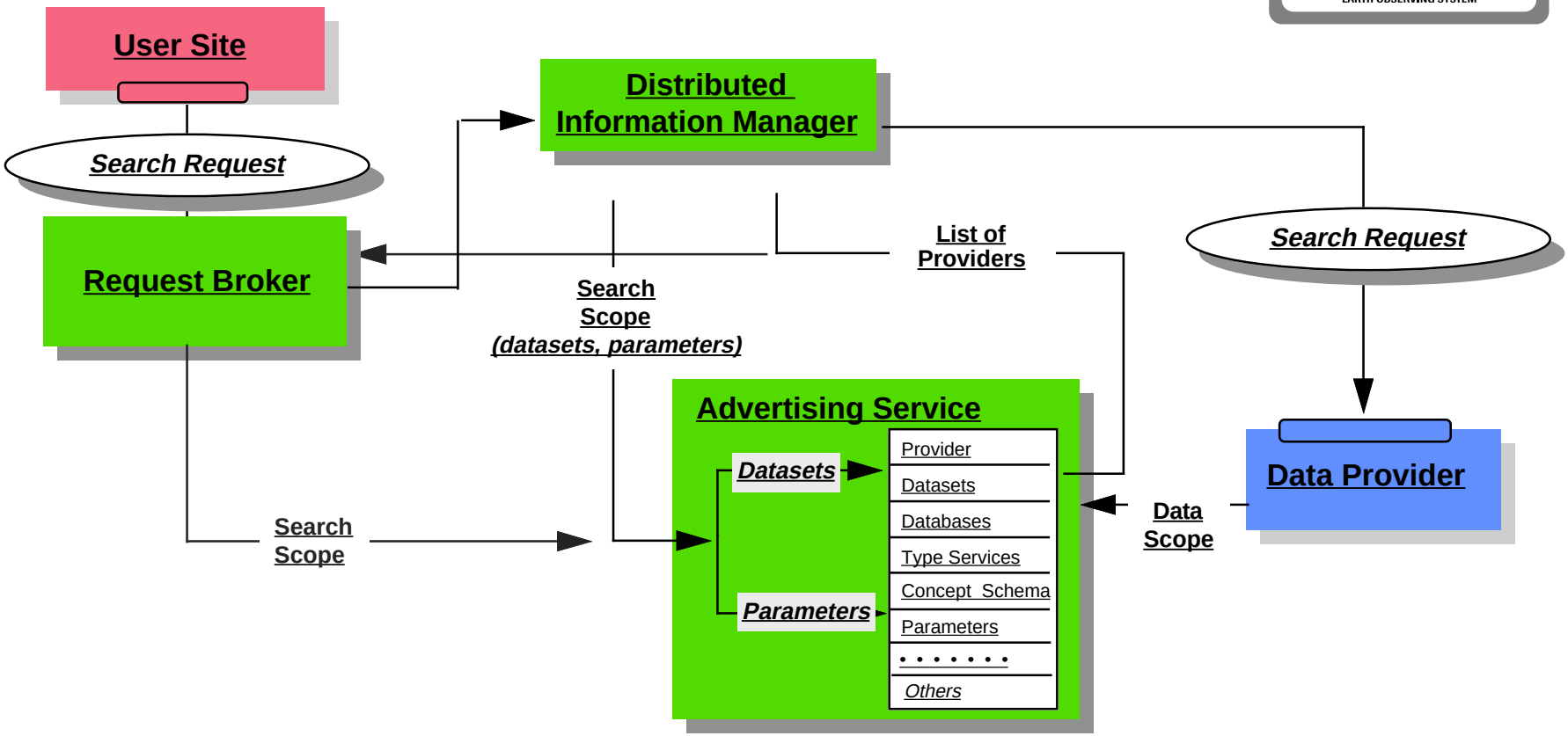
Inter-Site Search Vision

What is the distribution of Phytoplankton Concentration underneath the ozone hole in the ice-free ocean for dates nearest October 15, 1995



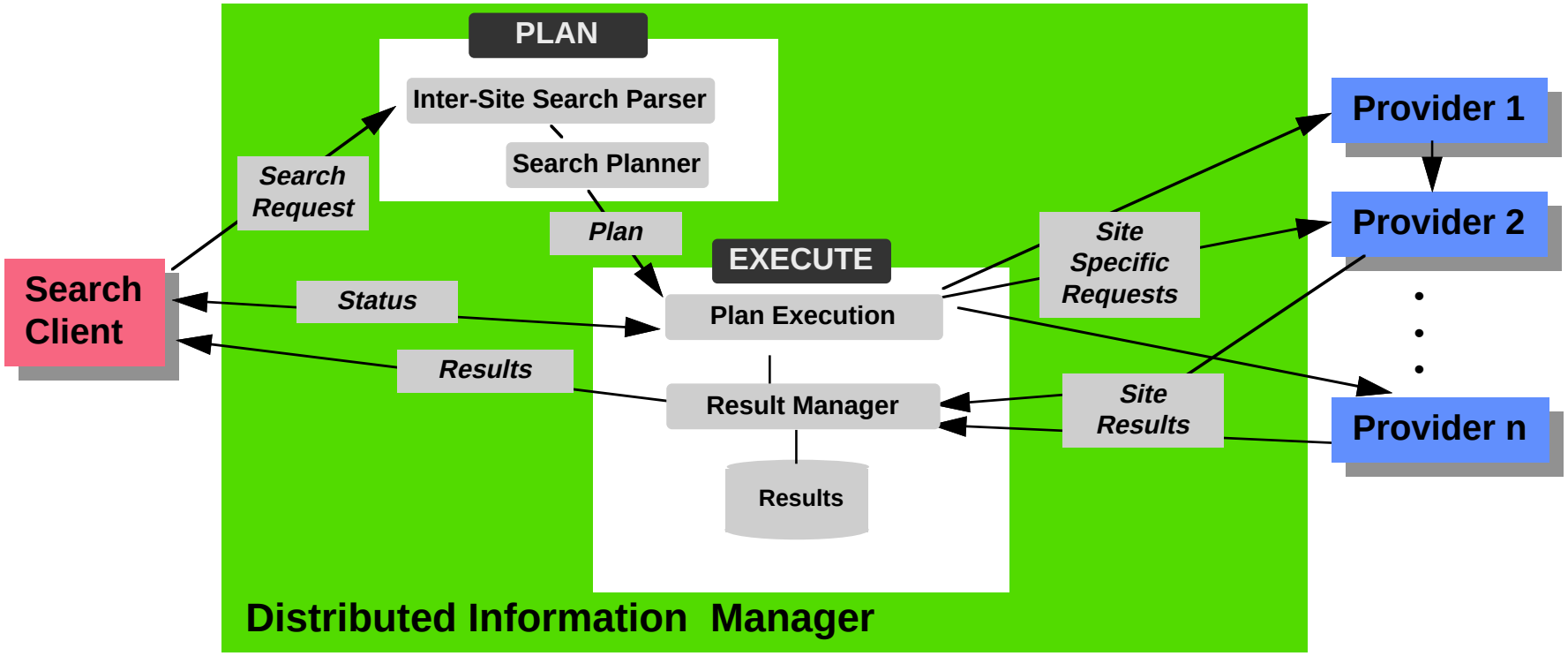
- Assign Operations to "Best Site" (i.e., optimization)
- Route Data/Results Directly Between Sites
- Formatted Results Returned to Requestor

Inter-site Search Routing



- Search Service Determines Best Performance Node
- Search Node Takes Responsibility For Search and Notification
- Data Providers Describe Data Scope In Advertisements

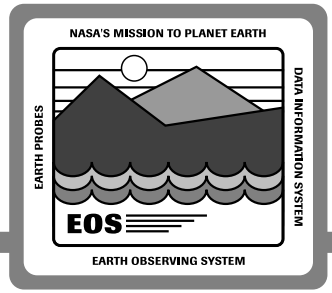
Inter-site Data Searching
Complex/Coincidence Search



Joins Referencing Event/Time/Space

Create Plan For Doing Joins, Unions, ...
Can Perform Joins, Unions, ...

Intersite Search Routing



key issues

- **context mapping (vocabulary)**
- **query decomposition**
- **optimization of mapping sub-requests**

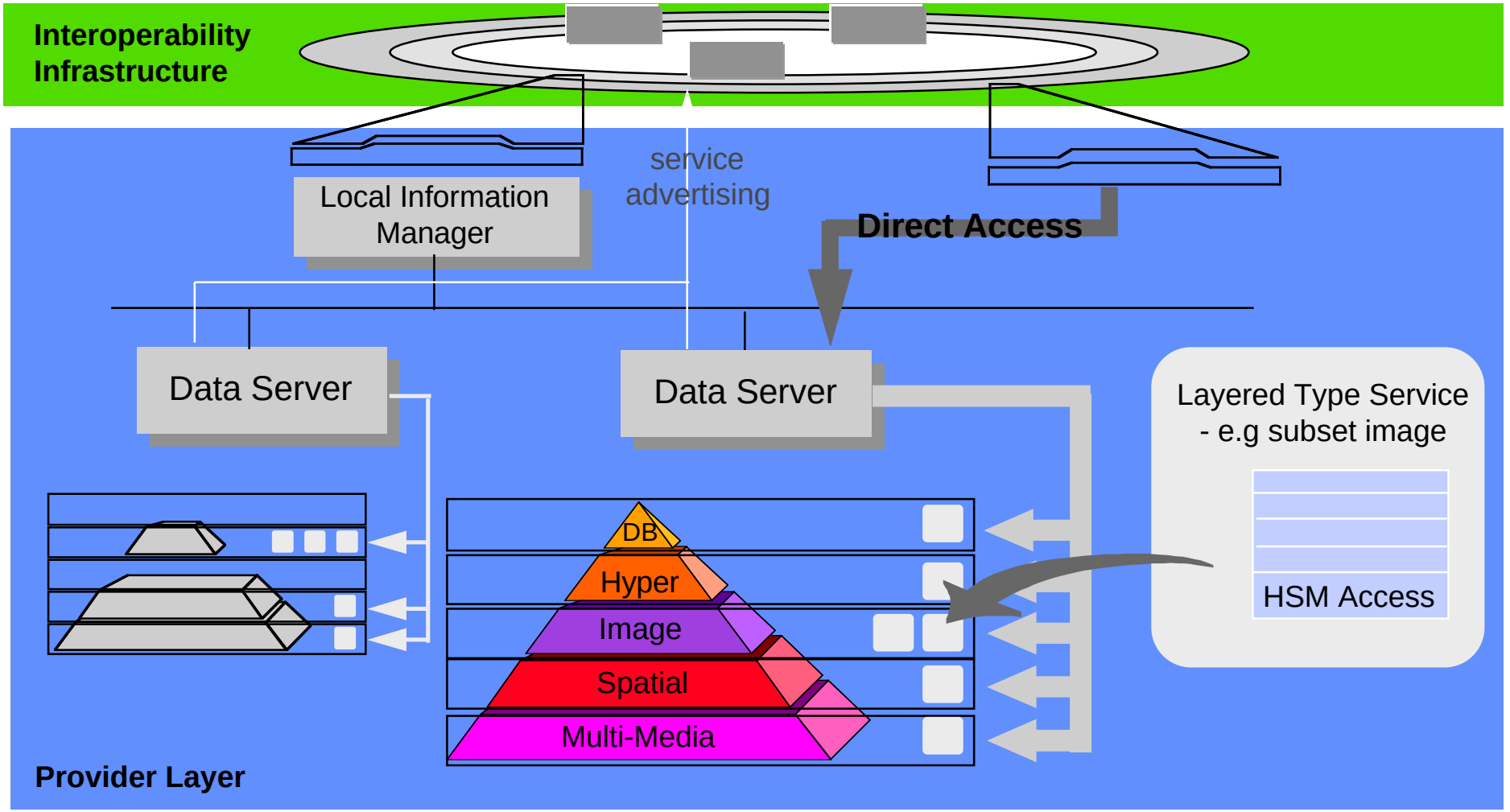
current trade study

- **“service routing” investigates these issues**

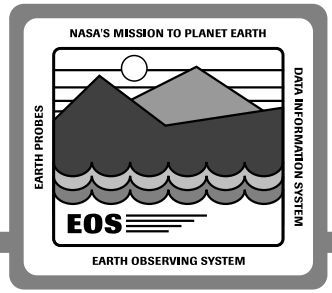
incremental development steps

- **establish infrastructure with appropriate architecture direction**
- **develop increasingly sophisticated components**
- **encourage third-party R&D and implementation of components**

Data Servers



Supports Content-based Search



features

- data servers - provide access (DIM or direct) to collection of related data for which a provider wants to support through services - there is no pre-conception of data collection structure (can support complex items)
- all data collections are not equivalent
- data servers capabilities are advertised to the network by providers using an extensible service description language
- data set servers built on re-usable “type” services (rdbms, text, image, etc.) - layered components which link to data through the HSM
- type services should support polymorphism
- multiple services to the same layer of data - not all ECS developed (DAAC, external)

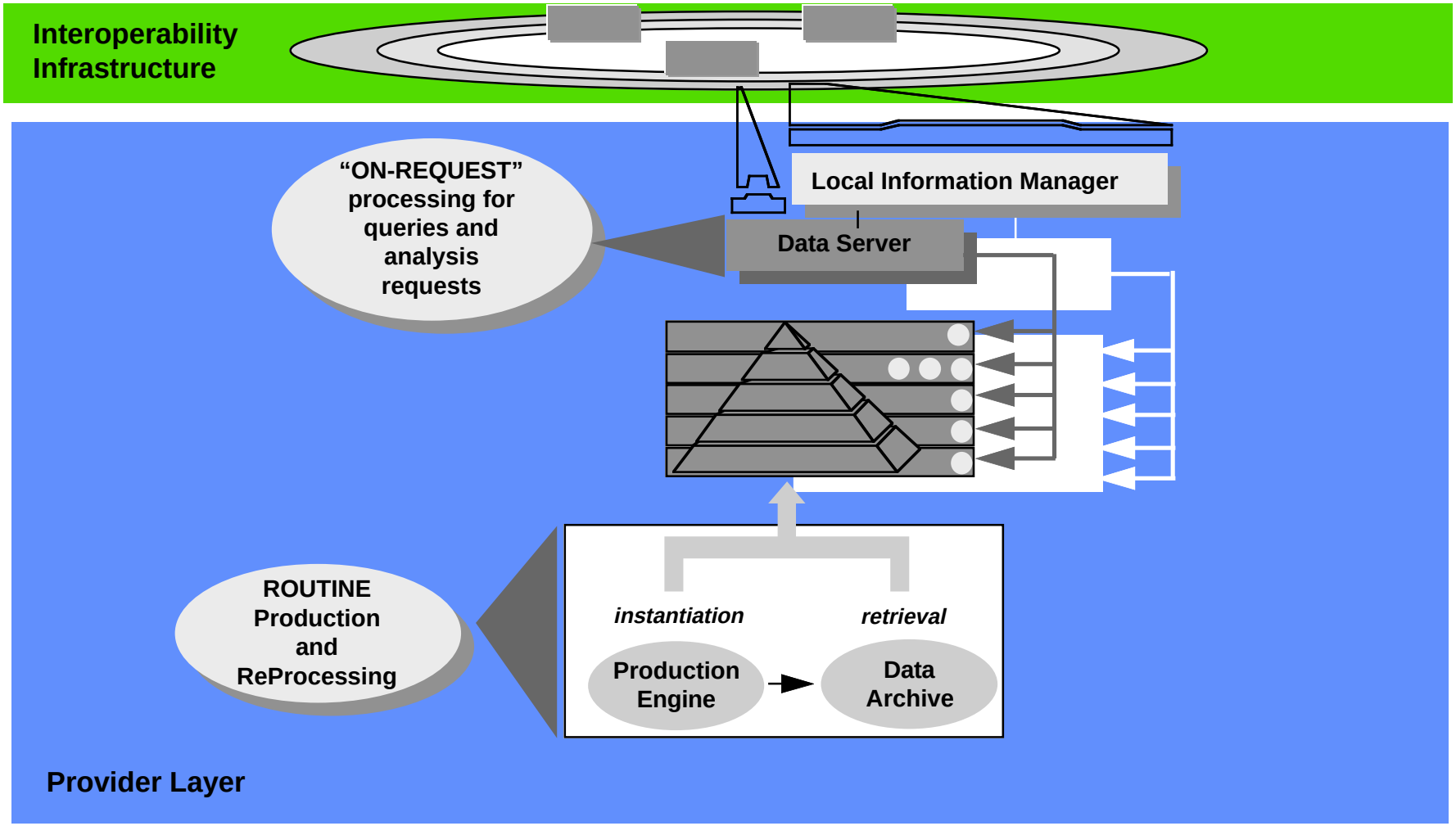
issues

- language extensions to support content-based search - *current trade study*
- support for user-method insertion as type services

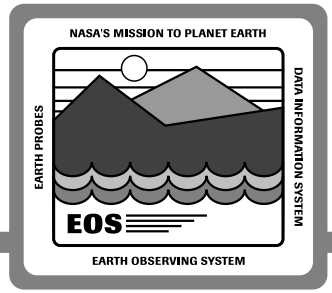
incremental steps

- COTS isn’t there yet, but research is underway (SQL3, OO data languages, etc.)
- need short / long term strategy - with increasing functionality

Data Access


Interoperability Infrastructure
"ON-REQUEST" processing for queries and analysis requests
Local Information Manager
Data Server
ROUTINE Production and ReProcessing
instantiation
retrieval
Production Engine
Data Archive
Provider Layer

Data access



features

- **"old" distinction of PGS, IMS, DADS no longer valid - processing is required to meet user needs in all areas**
- **some sites will include extensive routine production/re-processing**
- **logical use of same processing capacity for "on-request" processing**

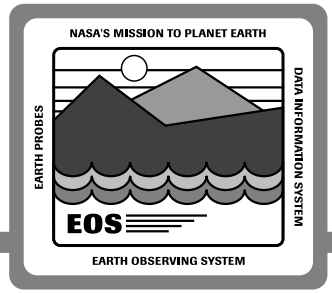
issues

- **referencing data items, services**
- **HSM support (COTS capability?) - *current trade study***
- **resource management for request processing**

incremental steps

- **develop universal referencing scheme which supports extension**
- **provide a uniform HSM access layer**

Status



- **attempting to architect around user needs**
- **there is still considerable detail to be added and issues to be tackled**
- **proposed architecture should allow progression towards the vision in an evolutionary manner**
    - **Near-Term**
        - **establish the infrastructure and basic components**
        - **adapt V0 components to work within infrastructure**
    - **Longer-Term**
        - **extend/replace the query language**
        - **improve the search functionality - including content-based search**
        - **improve DIM capability - allowing better optimization of requests among logically distributed service providers**

continuing development of type services to meet user needs and technology capabilities